Year 2 Report for NASA AISRP Grant NNG05GA94G

Project Title:
A Neural Map View of Planetary Spectral Images for Precision Data Mining and Rapid Resource Identification

Erzsébet Merényi, PI, Rice University
Robert H. Brown Co-I, University of Arizona, Tucson, AZ
William H. Farrand, Co-I, Space Science Institute, Boulder, CO
Thomas Villmann, Collaborator, Klinik für Psychotherapie, Universität Lepzig, Germany
Colin Fyfe, Collaborator, University of Paisley, Scotland

Project period: 11/1/2004 – 10/31/2007
Project URL: http://www.ece.rice.edu/∼erzsebet/HYPEREYE.html

Contents

1 Project Summary (from original proposal)

This project follows up on a current three-year AISRP grant, NAG5-10432, which will end 8/2004. It addresses a pressing need for rapid yet intelligent analysis of voluminous multi- and hyperspectral images in order to extract key data and generate knowledge. Spectral imaging plays a leading role in remote identification of surface materials of Earth (Landsat, AVIRIS, Hyperion), Mars (Pathfinder, MGS, MER), the Jovian system (Galileo NIMS), the Saturnian system (Cassini VIMS) and other solar system bodies. Hyperspectral sensors, in particular, enable detailed identification through the complexity of signatures measured in hundreds of narrow spectral bands. The challenge in automated and fast (real time, on-board) interpretation of these huge images calls for massively parallel algorithms, as well as it requires sophisticated algorithms for optimal knowledge extraction. Properly utilized, Artificial Neural Networks (ANNs) can provide both.

The current project engineered ANNs, specifically Self-Organizing Maps (SOMs) and their hybrids for efficient and sophisticated clustering and classification of spectral images, developing custom modules supported by commercially available components. Based on some of the latest theoretical research on SOMs, the tools we developed, jointly with European experts, are powerful for distinguishing a large number of spectral classes and for the discovery of "interesting" but uncommon and spatially very small classes. We use information theoretically principled SOM approaches, which increases power and confidence in autonomous data mining. We demonstrated the effectiveness and high quality of data analysis on sample IMP spectral images, Cassini VIMS Jupiter fly-by imagery, AVIRIS and other data representing typical challenges in NASA's missions. We propose to advance these computational intelligence capabilities in three ways: 1) We will add significantly new theoretical strength to information extraction modules. 2) The software, HYPEREYE, will be made transferable to other users through high-level graphic interface, augmented software design, tutorials and wrapping, opening an important phase of technology infusion that will take recent and future developments into the user community. 3) We will directly participate, using our methods and software, in analyses of spectral images forthcoming from the Mars Exploration Rovers and Cassini VIMS Saturn orbital tour, and (pending its funding) a Pluto/icy satellites spectral analysis project.

The 'neural' core of our software is already suitable for implementation in high-speed massively parallel hardware (which could be an on-board analysis capability), as it was one of the original objectives of our work. We are pursuing that line of development outside of this project proposal and, if successful, we anticipate using the hardware to support this work as well.

This project is a collaboration between computer science and space science investigators at Rice University, University of Arizona, and the Space Science Institute, Boulder, CO.

2 Accomplishments in Year 2, 11/1/05 - 10/31/06

2.1 Completion of project tasks (algorithm and software development, data analysis)

Task 1: Work on the main GUI, data handling and other support layers Expansion of data handling capabilities, started in Year 1, continued. Conversion of input data from a variety of binary and ascii formats to a unified internal binary format that can also contain meta data (done mostly in Year 1), was complemented by post-processing scripts for various interpretations of very large amounts of outputs from long clustering and classification runs. These serve both the summarization of scientific results for lists, tables, graphs, etc., and the evaluation of algorithm behavior. The meta data extension allows useful visualizations and annotations of data that have no spatial context. With the comprehensive ascii and meta data format in place we are now in the process of implementing the rather specialized visualization that uses meta data. We have been using the ascii / meta data facilities to process planetary ice spectra for a joint project (AISRP NNG05GA63G, PI Eliot Young.)

We continued to work on the update of HyperEye modules to use QT, which was originally planned for Year 2 but we started it in Year 1. Remodeling with QT is a) a forward looking act to prepare our software for easy porting to multiple platforms in order to facilitate technology transfer; (the actual porting to multiple platforms is beyond this project), b) is providing a support layer for updating/extending various existing functionalities. The QT library's visualization components aid work in Task 3 and elsewhere, assist meta data parsing, and exporting of results to various graphic formats. Later it will also help perform history recording of interactive user input during module operation. (We moved this history recording to Year 2 and 3 to gain efficiency by completing the above, more essential items first). We have now, by contribution of undergraduate student Allen Geer, a visualization library (Vismod), built on QT, which provides a unified frame for displaying all data that we handle, including neural net components / layers, as well as user data, and has the capability of using meta data for customized annotations. Using Vismod capabilities, we extended the plug-in architecture for visualization (developed in Year 1). Plug-ins are slaves to main modules. Development of customized visualization through plug-ins makes recompilation of the main module unnecessary when new capabilities are added to the plug-in, and a module can receive services from multiple plug-ins. This facilitates fast development of a variety of customized visualizations that can be used on demand by multiple modules. Our algorithm developments under Tasks 2 and 3 were greatly accelerated by the implementation and use if plug-ins.

Web demonstration of our main tools is at http://www.ece.rice.edu/~erzsebet/HYPEREYE.html, complete with online documentation, demo data sets and tutorial.

Task 2: Advancement of ANN/knowledge extraction algorithms: We further studied the capabilities of the SOM magnification algorithm (BDH) by Bauer, Der and Herrmann (1996). The BDH allows control of the magnification exponent α in the power law $Q(w) = const * P(v)^{\alpha}$, which is proven to express the relationship of $Q(w)$, the density of SOM weight vectors (the quantization prototype vectors) in input data space, to $P(v)$, the density of the input data. $\alpha = 1$ therefore provides the information theoretically optimal (max. entropy) SOM mapping, and $\alpha < 0$ enlarges the SOM response areas for rarely occurring data, thereby increasing the chance of their discovery. Since none of the conventional quantization algorithms produces such magnification values, the BHD is potentially very powerful for manifold learning and sophisticated data mining. However, since the BDH theory can only provide analytical proof for 1-d data and for n-d data whose dimensions are statistically independent, we must investigate its limits for more complex data with numerical simulations. We showed the knowledge discovery power of SOM magnification on moderate-dimensional spectral images of Mars and Earth. Our paper submitted last year to IEEE Trans. Neural Networks (Merényi, Jain, Villmann (2005)) was revised and accepted. We continued this investigation, moving toward hyperspectral dimensions.

In March, 2005, graduate student Lily Zhang started an investigation of the Topographic Function (TF, Villmann *et al.*, 1997), a measure of the topology preservation in SOM mapping (of any kind, thus BDH too), to determine the TF's suitability for high-dimensional data. Topology preservation is a prerequisite in any mapping method that aims at correct cluster discovery from the lower-dimensional "projection" of a higher-dimensional data space. Lily's work resulted in an *original measure, the Weighted Differential Topographic Function (WDTF)*. The WDTF has been demonstrated to reveal more sophisticated facts about the mapping quality for complicated data (such as hyperspectral imagery), and was presented at the European Symposium of Artificial Neural Networks, 2006 (Zhang and Merényi, 2006). We are working on an extended wersion for a journal publication. Lily implemented this new WDTF algorithm along with the original TF as a HyperEye module (called TF).

Graduate student Major Michael Mendenhall has been working with the PI on "relevance learning" with Generalized Relevance Learning Vector Quantization (following Hammer and Villmann, 2002; Villmann *et al.*, 2003) to assess the relative importances of hyperspectral data dimensions. We investigate the GRLVQ because we have not seen good feature extraction for hyperspectral data, *i.e.*, schemes that preserve the distinction among the many classes that hyperspectral data is meant to distinguish. The GRLVQ is the first non-parametric method we know of, to measure the quality of the feature selection by the classification accuracy on the same data. Non-parametric approach is important for hyperspectral data where data models and priors are usually unavailable. We found an instability in the Hammer and Villmann scheme, devised and implemented a modified learning (GRLVQ-Improved, or GRLVQI) that stabilized the process and increased the classification accuracy and speed.

A *fundamental finding* of this research is that when GRLVQ is applied in the wavelet domain the wavelet coefficients that are selected as the N most relevant ones for a given classification, are not the same as the N largest magnitude coefficients! While this may be surprising, and is certainly different from the prevailing practice of wavelet coefficient selection for best modeling of data, it can be explained: The N *largest wavelet coefficients* ensure *minimum distortion* restoration of the data from the quantized encoding. In contrast, the coefficients with the N *largest relevances* provide *maximum information* relative to the class discrimination requirements of the given task. We achieve higher classification accuracy with a substantially reduced set of input features (in both the original reflectance data domain and in the wavelet transform domain), than with the full input. This could not be achieved with traditional feature extraction approaches such as PCA, or conventional selection of wavelet coefficients.

We presented a paper at the Annual Meeting of the Am. Soc. of Photogrammetry and Remote Sensing (Mendenhall and Merényi, 2006a), submitted a paper to IEEE Trans. Neural Networks (Mendenhall and Merényi, 2006b), presented a paper at the 2006 IEEE Mountain Workshop on Adaptive and Learning Systems, SMCal 2006 (Mendenhall and Merényi, 2006c), and are preparing a paper for submission to IEEE Trans. Geoscience and Remote Sensing. Major Mendenhall is supported by the U.S. Air Force, his contributions have been at no cost to this grant except support for the SMCal 2006 workshop.

Dr. Villmann visited in March, 2005, during which time, and at ESANN 2006 in April 2006, we had intense discussions on all three topics above. He is on Major Mendenhall's PhD committee.

Investigation of the GSOM (Growing Self-Organizing Map), originally planned for Year 2, was postponed in favor of the exciting developments from our research in TF and GRLVQ described above.

Task 3: Improvement of visualization and human interaction Last year graduate student Kadim Taşdemir and the PI developed a highly automated method for cluster boundary extraction from converged SOMs (by clustering the SOM prototypes). The corresponding HyperEye module, SOMcluster, seems more sensitive to subtleties of data structures and is more automated than current methods in the published literature. Fully automated solution does not exist to date. We regard our previous results, presented at the highly selective Workshop on Self-Organizing Maps, Paris, Sep 5-8, 2005 (Taşdemir and Merényi, 2005), as an intermediate stage toward precise and automated cluster boundary detection, and continued to work toward that goal this year. One focus has been the choice of inter- and intra-cluster distances, as well as the choice of cluster validity index, and the possible need for new measures because the complexity of the data of interest is not always handled properly by existing measures.

Taşdemir and the PI proposed an *original visualization* of data manifolds, draped over an SOM (or potentially over any other mapping scheme). This representation uses a weighted version of the so-called (binary) Adjacency Matrix produced by Delaunay triangulation (a fairly standard way to represent a manifold with a given set of quantization prototypes), where the weighting of the connection between two Voronoi centroids is the number of data points for which one centroid is the closest, the other is the second closest prototype. Hence the elements of this weighted adjacency matrix, which we call *Connectivity Matrix (CONN)*, express not only the connectedness of the manifold, but also the local density, i.e., how strongly various regions are connected. This greatly facilitates cluster capture in a noisy data set, including the identification of outliers. We presented this method at the European Symposium of Artificial Neural Networks, 2006 (Taşdemir and Merényi, 2006), and are working on an extended version for a journal publication. Kadim implemented this visualization algorithm as a HyperEye plug-in (see Task 1), CONNvis, with interactive user-driven query features.

Graduate student Lily Zhang created a plug-in (see Task 1), TopoView, for flexible visualization and evaluation of topology violations. The software units CONNvis, TopoView, and TF are envisioned as components of a future unified tool to monitor the quality of SOM mapping during learning.

Last year, Allen Geer, an undergraduate student wrote a sophisticated plotting package (HyperEye module VECPLOT) that has the intelligence to arrange plots such that all graphs show appropriately (for example, properly offset for viewing), class labels or legends are placed where they should be, etc., yielding great savings of time and effort in our data analyses. This year, before he graduated and left Rice, Allen completed a new library called Vismod (built on QT, see Task 1) to provide a unified scheme for graph and image display, and user interaction needs of our modules. Our plug-ins (above) and some of our new modules are already using the Vismod library. We also started updating older, major modules with Vismod/QT. The complexity of some of our major modules, namely that of our data exploration module SPECTER, however, presented some hard surprises that necessitate thorough analysis of the components' interaction, and re-evaluation of this approach, before we can confidently proceed. Allen's work was supported by internships from Rice U School of Engineering, and from the PI's startup funds.

Task 4: Applcation of HyperEye algorithms to real scientific data Last year we clustered and classified IMP SuperPan spectral images. This turned out more difficult than originally expected, because of calibration issues including a mosaic effect that manifests in different viewing geometry and/or instrument characteristics for various patches within each image. Due to this, signatures of the same material class can be sufficiently different across octants that a classifier trained on one octant will not recognize the same materials in another octant. Only for a few octants did we obtain good enough supervised class maps based on training with prototypes from one reference octant. Those were interpreted, compared to previous analyses, and presented at LPSC (Farrand *et al.*, 2005). Additional work was done this year, presented at the AGU fall meeting and published in Wright *et al.*, (2005). We are still working on a journal paper for Icarus (Farrand *et al.*, 2005b). Ours is the first comprehensive classification of the SuperPan images acquired in 1997, due to the above difficulties. To cope with the above "mosaic" problem we would need to do full clustering on every octant, label the clusters, perform supervised classification, and then find the corresponding classes across all octants. Cross-matching clusters or classes across multiple images is a tedious, labor intensive, and often difficult task requiring external knowledge. In general it would require some automated semantic labeling, which is outside of the scope of this project. We did manual cross-matching for two octants but do not have resources to do more.

Co-I Farrand provided MER images from both the Spirit and the Opportunity rovers. These data are very similar to the IMP data in terms of structure, but are much cleaner and without the mosaic effect. Clustering and classification is in progress. We expect to submit preliminary results to the next LPSC. We have not received Cassini VIMS data so far, for reasons of priority of other pressing urgencies for the VIMS PI (R. Brown) and team. We plan VIMS analysis for the reminder of Year 2, and for Year 3, as we receive data. Analysis on AVIRIS terrestrial hyperspectral images to map clay-bearing surface units, in support of a landslide hazard assessment project (NASA SE&NH, V. Baker PI, U AZ) produced excellent results last year, reported in Rudd and Merényi, 2005. This investigation was a large contribution to Larry Rudd's PhD Thesis, and we are now preparing an article for JGR. All "relevance learning" work with graduate student Mike Mendenhall (described in Task 2) was also done on terrestrial AVIRIS hyperspectral imagery (of the Lunar Crater Volcanic Field). Work on terrestrial analogs with well known ground truth is a very important component of our development of computational intelligence algorithms for planetary applications. This project also supported joint work with AISRP project NNG05GA63G (PI Eliot Young), for which we have been developing neural classifier models for the prediction of surface temperature and grain size of surfaces of distant planetary bodies (such as Pluto) from ice spectra.

2.2 Publications, presentations, other related activities in this award year

Refereed journal and refereed conference proceedings

Mendenhall, M.J., Merényi, E., (2006c), Relevance-based Feature Extraction from Hyperspectral Images in the Complex Wavelet Domain, Proc. IEEE Mountain Workshop on Adaptive and Learning Systems (SMCals/06), Logan, Utah, July 24 – 26, 2006. pp. 24–29

Mendenhall, M.J., Merényi, E. (2006b), Relevance-based Feature Extraction for Hyperspectral Images, Submitted to IEEE Trans. Neural Networks.

Mendenhall, M.J., Merényi, E., (2006a), Generalized Relevance Learning Vector Quantization for Classification Driven Feature Extraction from Hyperspectral Data, Proc. ASPRS 2006 Annual Conference and Technology Exhibition, Reno, NV, May 1-5, 2006.

Taşdemir, K., Merényi, E., (2006), Data topology visualization for the Self-Organizing Map, Proc. 14th European Symposium on Artificial Neural Networks, ESANN 2006, Bruges, Belgium, 26–28 April, 2006. pp. 125–130

Zhang, L., Merényi, E., (2006), Weighted Differential Topographic Function: A Refinement of the Topographic Function, Proc. 14th European Symposium on Artificial Neural Networks, ESANN 2006, Bruges, Belgium, 26–28 April, 2006. pp. 13–18

Merényi, E., Jain, A., Villmann, Th. (2005), Explicit Magnification Control of Self-Organizing Maps for "Forbidden Data", Accepted to IEEE Trans. Neural Networks.

Wright, S. P, Farrand, W. H., Rogers, A. D., Merényi, E., (2005), The Nature of the Mars Pathfinder "Black Rock" Lithology: Comparisons with SNC Meteorites and OMEGA Spectral Images of Chryse Planitia., Eos Trans. AGU Fall Meet. Suppl., Abstract P21B-0145. San Francisco, CA, December, 2005. 86(52)

Farrand, W. H., Merényi, E., Bell, J. III., Johnson, J., Murchie, S. and Barnouin-Jha, O. (2005b), Class maps of the Mars Pathfinder landing site derived from the IMP SuperPan: Trends in rock distribution, coatings and far field layering, In preparation.

Recent honors, PI

• Member of Editorial Advisory Board, International Journal of Intelligent Computing in Medical Science and Image Processing (IC-MED Journal), from 2006 (new journal, first issue expected in 2007)

• Associate Editor, Neurocomputing, 2006 - 2007

• "By Invitation only" participation in Dagstuhl Seminar "Similarity-based Clustering and its Application to Medicine and Biology, Int'l Conference and Research Center for Computer Science, March 25-30, 2007, Schloss Dagstuhl, Wadern, Germany.

Presentations by PI and team members

Relevance-based Feature Extraction from Hyperspectral Images in the Complex Wavelet Domain. IEEE Mountain Workshop on Adaptive and Learning Systems (SMCals/06), Logan, Utah, July 24 - 26, 2006. (Mendenhall, M.J., and Merényi, E.)

Generalized Relevance Learning Vector Quantization for Classification Driven Feature Extraction from Hyperspectral Data, Proc. ASPRS 2006 Annual Conference and Technology Exhibition, Reno, NV May 1-5 2006. (Mendenhall, M.J., and Merényi, E.)

Data topology visualization for the Self-Organizing Map. Proc. 14th European Symposium on Artificial Neural Networks, ESANN'2006, Bruges, Belgium, 26-28 April, 2006. (Taşdemir, K. and Merényi, E.)

Weighted Differential Topographic Function: A Refinement of the Topographic Function. Proc. 14th European Symposium on Artificial Neural Networks, ESANN'2006, Bruges, Belgium, 26-28 April, 2006. (Zhang, L. and Merényi, E.)

The Nature of the Mars Pathfinder "Black Rock" Lithology: Comparisons with SNC Meteorites and Omega Spectral Images of Chryse Planitia. Proc. Am. Geophys. Union fall conference, San Francisco, CA, December, 2005. (Wright, S. P, Farrand, W. H., Rogers, A. D., Merényi, E. (2005)

Finding Interesting Structures in High-Dimensional Data through Innovative Visualization. Industrial Affiliates Day, Electrical and Computer Eng. Dept., Rice University, September 7, 2006 (Taşdemir, K. and Merényi, E.)

International conferences the PI served on

• Track Co-Chair, World Automation Congress 2006 (WAC2006) (track: Int'l Forum on Multimedia and Image Processing), Budapest, Aug 24 - 26, 2006.

International program committees:

• IASTED International Conference on Computational and Systems Biology (CASB 2006), Dallas, USA, November 13-15, 2006

• 14th European Symposium on Artificial Neural Networks, ESANN'2005, Bruges, Belgium, April 26 - 28, 2006

• 6th Jordanian International Electrical & Electronics Engineering Conference (JIEEEC) 2005, November 15 - 17, 2005, Amman, Jordan

2.3 Proposals granted, renewed or submitted in this period (excluding this grant)

• Year 2 renewal of "Neural-Net Analysis of Icy Volatiles in the Solar System" PI Eliot Young, SWRI, Boulder, CO. NASA AISRP, NNG05GA63G, 10/1/2004 to 09/30/2007, $371,367 total budget, $45K/year to Co-I Merényi. Granted.

• " Adaptive Algorithms For Optimal Classification and Compression of Hyperspectral Images", NASA AISRP. PI Tamal Bose, Utah State U. $458,210 total; $168,945 subcontract to Rice U, for 3/01/06 - 2/28/08. Granted.

• " Utilizing spectral stratigraphy for improving mapping and understanding of layered rock sequences: Terrestrial examples with relevance for Mars ", NASA Mars Fundamental Research; Program. PI W.H. Farrand, Space Science Institute, Boulder, CO. $312,121 total; approx. $65,000 subcontract to Rice U, for 4/1/2007 - 3/31/2010. Pending.

• Super Spectrum Visualization Initiative (SSVI), A Solicited White Paper for NEXIS, Defense Intelligence Agency. PI Sean Anklam, SpecTIR LLC, Sparks, Nevada. Approx. $100K total, for 6 months, $13.5K subcontract to Rice U. Pending.

3 Work Plan for Year 3, 11/1/06 - 10/31/07

Task 1, Work on the main GUI, data handling and other support layers (~25 - 30% effort)

The update of HyperEye modules to use QT and our custom software support components (e.g., Vismod) will continue. History recording of interactive user actions during module operation will be added to these updated modules, postponed from Year 2.

Implementation of a new concept (that we developed in Year 2), a "data document object", will provide for a more unified and more flexible treatment of data services to modules for all internal and user data (i.e., descriptors of neural network properties as well as data to be analyzed, meta data, etc.), than we had so far and will accelerate the integration of modules with the main GUI in a cleanly structured manner.

During Year 3, the need for expansion of data formats/handling is mainly anticipated in the way of customized post-processing capabilities as scientific results produced by neural processing become more abundant and richer. We will add post-processing capabilities as demanded by the scientific output.

Our website for public demonstration will be updated, but instead of a java-based viewer planned earlier we want to consider python or similar facility to ensure necessary performance. Our current web demo is easy to view from unix machines, but it requires a cumbersome communications set up from PCs. Interested Co-Is (with compatible Sun/unix platforms and necessary third party components) can receive module releases in Year 3.

Task 2, Advancement of ANN/knowledge extraction algorithms (~25–30% effort): We continue, as planned, systematic evaluation of the scope and power of SOM magnification. We will further examine our topology preservation measure (WDTF) that improved on the Villmann topology preservation measure for SOMs. We have plans to develop an entirely new concept, the "cluster topographic function" (CTF), which measures topology preservation "on the cluster level". This is in response to the observation that a mapping does not have to be perfectly topology preserving in order to show correct cluster structure, it can be "scrambled" within data clusters (but not scrambled across clusters). We propose to create and implement a measure that distinguishes these levels of topology preservation. Existing measures tend to indicate high levels of topology violations when in fact a(n SOM) map is already good enough for correct cluster capture. A preliminary hint to this was published in Zhang and Merényi (2006). We expect that eventually, the CTF can be a monitoring and feedback tool for guiding the learning of complicated, large data manifolds.

GRLVQ research for the assessment of the relative importance of hyperspectral data dimensions continues. After the fresh results obtained in Year 2 we want to mature and refine the tools.

We may start systematic evaluation of the Growing Self-Organizing Map (GSOM) by Bauer and Villmann (1997), postponed from Year 2, depending on the success of and interest in, the above topics.

Visits by our German a Scottish collaborators, Drs. Villmann and Fyfe are planned for fall, 2006 and spring, 2007.

Task 3, Improvement of visualization and human interaction remains a substantial effort ($\sim$ 15 - 20%). The PI and Kadim Taşdemir continue research on automation of cluster boundary extraction from learned SOMs. As stated in Task 3 accomplishments, implementation of existing, and possibly invention of new cluster distance measures and cluster validity indices, will remain one important aspect.

Exploitation of QT/Vismod-based support is on-going, and as described last year. Various visualizations are being turned into plug-ins (for example the rendering of SOM knowledge, the variants of which are now wired into the SOM learning module), so they can be easily used by any module. With Vismod, the foundation of visualization of neural network "objects", data objects and related meta data (annotations) already exists. Building custom visualization and/or tuning existing capabilities and remodeling the data processing modules to use them, will go on in Year 3 too, as planned.

The plug-ins CONNvis, TopoView, and TF (described in Tasks 2 and 3) will be further developed into a unified tool to monitor the quality of SOM mapping during learning. We expect that this monitoring will also collaborate with automated cluster capture.

Task 4, Application of HyperEye to real scientific data will be $\sim$ 25–30% of the effort. We will finish data analysis from Pathfinder, and shift to the Mars Exploration Rovers, for Mars data analysis. We expect to work with Cassini VIMS data. Terrestrial hyperspectral data will be analyzed to more comprehensively evaluate GRLVQ(I). We continue to work on Eliot Young's planetary ice spectra (in a joint project), which produced very encouraging results in Year 2.

As a new component, we plan to create one or two "reality based synthetic" spectral images, which will combine the characteristic real properties of existing, large and complicated spectral images with the convenience of a completely labeled data set. We feel a tremendous need for building such realistic synthetic data since the data sets in standard national repositories do not represent the sophistication of planetary spectral data and thus do not pose sufficient challenge for our algorithms. At the same time, for rigorous evaluation of our algorithms we must have complete labeling for appropriately sophisticated test data sets, which does not exist for (hyper)spectral imagery. The details of generating these data are non-trivial, but if we succeed, we can offer the resulting data for an AISRP "challenge" repository.

We expect journal submissions in Year 3 on topology preservation measures (Task 2), CONN visualization and clustering (Task 3), and Mars Pathfinder and terrestrial hyperspectral image classification (Task 4). Potential conference submissions include new topology preservation measure (CTF) and MER image analysis.

4 References (other than listed under section 2.2)

Some of the PI's papers are downloadable at http://www.ece.rice.edu/~erzsebet/publications.html

Bauer, H.U., Der. R., Herrmann, M. (1996), Controlling the Magnification Factor of Self-Organizing Feature Maps, Neural Computation, 8:757–771.

Bauer, H.U., and Villmann, Th. (1997), Growing a Hypercubical Output Space in a Self-Organizing Feature Map, IEEE Trans. on Neural Networks, 8(2):218-226.

Farrand, W. H., **Merényi, E.**, Murchie, S., Barnouin-Jha, O. (2005), Spectral Class Distinctions Observed in the MPF IMP SuperPan Using a Self-Organizing Map., Proc. 36th Lunar and Planetary Science Conference, Houston, Texas, March, 2005. , (Extended abstract).

Farrand, W. H., **Merényi, E.**, Bell, J. III, Johnson, J., Murchie, S. and Barnouin-Jha, O (2005b), Class maps of the Mars Pathfinder landing site derived from the IMP SuperPan: Trends in rock distribution, coatings and far field layering, In preparation.

Hammer, B., Villmann, Th. (2002), Generalized relevance learning vector quantization, Neural Networks, 15:10593/41068.

Merényi, E., Jain, A., Villmann, Th. (2005), Explicit Magnification Control of Self-Organizing Maps for "Forbidden Data", Accepted to IEEE Trans. Neural Networks

Rudd, L., **Merényi, E.**, (2005), Assessing Debris-Flow Potential by Using AVIRIS Imagery to Map Surface Materials and Stratigraphy in Cataract Canyon, Utah, Proceedings of the Fourteenth Airborne Earth Science Workshop, Pasadena, CA, May 24-27, 2005 (Ed. R.O.Green)..

Taşdemir, K., **Merényi, E.**, (2005), Considering Topology in the Clustering of Self-Organizing Maps., Proc. 5th Workshop On Self-Organizing Maps (WSOM 2005), 5 - 8 September, 2005, Paris, France. Accepted. pp. 439-446

Villmann, Th., Herrmann, R.Der, and Martinetz, Th. (1997), Topology Preservation in Self-Organizing Feature Maps: Exact Definition and Measurement, IEEE Trans. on Neural Networks, 8(2):256–266.

Villmann, T.,**Merényi, E.**, Hammer, B. (2003), Neural Maps in Remote Sensing Image Analysis, Neural Networks, Special Issue on Neural Networks for Analysis of Complex Scientific Data, 16:(3–4):389-403.

5 Appendix: Acronyms

ANN	Artificial Neural Network
AVIRIS	Airborne Visible and Infrared Imaging Spectrometer, of NASA, JPL
AVIRISLA	Low-Altitude AVIRIS
BDH	algorithm for SOM magnification control by Bauer, Der and Herrmann (1996)
CONN	Connectivity Matrix
BP	Back Propagation, an ANN paradigm
EO-1	Earth-Observing satellite, NASA
GSOM	Growing Self-Organizing Map, an ANN paradigm
GRLVQ	Generalized Relevance Learning Vector Quantization
GRLVQI	Generalized Relevance Learning Vector Quantization Improved
HST	Hubble Space Telescope
HYDICE	Hyperspectral Digital Image Collection Experiment, Naval Research Lab
IMP	The Imager for Mars Pathfinder
Landsat TM	Landsat Thematic Mapper
Landsat MSS	Landsat Mulstispectral Scanner
LPL	Lunar and Planetary Laboratory; at University of Arizona
MER	Mars Exploration Rovers
MGS	Mars Global Surveyor
ML	Maximum Likelihood classifier
PCA	Principal Components Analysis
SOM	Self-Organizing Map, a neural network paradigm
STIS	Space Telescope Imaging Spectrograph
TF	Topographic Function (a topology preservation measure)
UA	University of Arizona
VIMS	Visible-Infrared Mapping Spectrometer, Cassini mission
WDTF	Weighted Differential Topographic Function (a topology preservation measure)

6 Budget for Year 3

Our budget, attached, for Year 3 remains as it was approved except for small increase in student stipends. Rice University's ECE Department adapted an aggressive policy to increase student stipends as part of the Department's recruitment strategy. Faculty is strongly urged to follow these new stipend guidelines. Our modified request reflects a mid-point between originally budgeted stipends and Department recommendation (for third and fourth year students such as Lily Zhang and Kadim Taşdemir on this project).

YEAR 3

Fiscal Year Ratios A= 0.6666667 B= 0.3333333

Percent of total year which is off campus	2007 FY-A Off-campus	2008 FY-B Off-campus	2007 FY-A On	2008 FY-B On	Total For Year 3
			13,162	6,581	19,742
			13,162	6,581	19,742
			26,038	13,019	39,057
			20,533	10,267	30,800
			59,733	29,866	89,599
			18,256	9,128	27,384
			77,989	38,994	116,983
				6,150	6,150
				7,000	7,000
			359	179	538
			800	400	1,200
			2,533	1,267	3,800
			8,007	4,004	12,011
			11,699	5,850	17,549
			89,688	57,994	147,682
			39,709	25,545	65,254
			129,397	83,539	212,936
			129,397	83,539	212,936
MTDC			**77,860**	**50,089**	**127,949**
GSTR			**7,803**	**3,901**	**11,704**
	Off-campus	Off-campus	On-campus	On-Campus	Total

SUMMARY PROPOSAL BUDGET Year 3

ORGANIZATION William Marsh Rice University			PROPOSAL NO.	DURATION (MONTHS)
PRINCIPAL INVESTIGATOR/PROJECT DIRECTOR Erzsebet Merenyi			AWARD NO.	

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with name and title. (A.7. Show number in brackets)	Person-months CAL	ACAD	SUMR	Funds Requested By Proposer	
1. Erzsebet Merenyi			1.70	$ 19,742	
2. Robert Brown					
3. William H. Farrand					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION PAGE)					
7. (3) TOTAL SENIOR PERSONNEL (1-6)			1.70	19,742	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. ( ) POST DOCTORAL ASSOCIATES					
2. (1) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	8.25			39,057	
3. (2) GRADUATE STUDENTS total of 14.0 months				30,800	
4. ( ) UNDERGRADUATE STUDENTS					
5. ( ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)					
6. ( ) OTHER					
TOTAL SALARIES AND WAGES (A+B)				89,599	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS) See the budget attachment for explanation				27,384	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A+B+C)				116,983	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT					
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)				6,150	
2. FOREIGN				7,000	
F. PARTICIPANT SUPPORT COSTS 1. STIPENDS $ 2. TRAVEL 3. SUBSISTENCE 4. OTHER					
TOTAL NUMBER OF PARTICIPANTS) TOTAL PARTICIPANT COSTS					
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				538	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				1,200	
3. CONSULTANT SERVICES					
4. COMPUTER SERVICES				3,800	
5. SUBAWARDS				12,011	
6. OTHER					
TOTAL OTHER DIRECT COSTS				17,549	
H. TOTAL DIRECT COSTS (A THROUGH G)				147,682	
I. INDIRECT COSTS (F&A) (SPECIFY RATE AND BASE) See Budget Attachment					
TOTAL INDIRECT COSTS (F&A)				65,254	
J. TOTAL DIRECT AND INDIRECT COSTS (H+I)				212,936	
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPG II.D.7.j.)					
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				$ 212,936	$

M. COST SHARING: PROPOSED LEVEL $	AGREED LEVEL IF DIFFERENT: $
PI/PD TYPED NAME & SIGNATURE* Erzsebet Merenyi	DATE
INST. REP. TYPED NAME & SIGNATURE* Jordan Konisky, Vice Provost for Research and Graduate Studies	DATE